

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



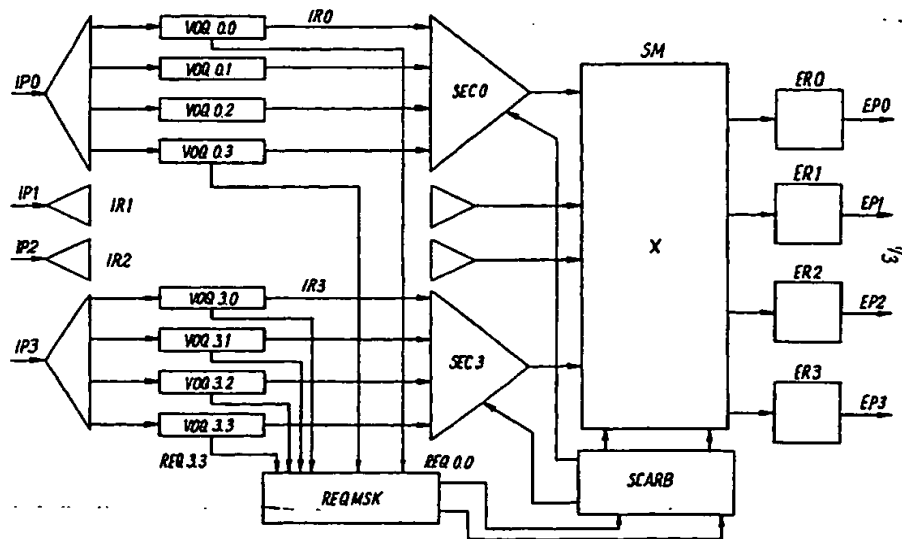
(43) International Publication Date
11 October 2001 (11.10.2001)

PCT

(10) International Publication Number
WO 01/76157 A1

- (51) International Patent Classification⁷: **H04L 12/56** (74) Agents: MCNEIGHT, David, Leslie et al.; McNeight & Lawrence, Regent House, Heaton Lane, Stockport, Cheshire SK4 1BS (GB).
- (21) International Application Number: PCT/GB01/01073
- (22) International Filing Date: 12 March 2001 (12.03.2001) (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 0008195.0 5 April 2000 (05.04.2000) GB (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- (71) Applicant (*for all designated States except US*): POWER X LIMITED [GB/GB]; Stafford Court, 145 Washway Road, Sale, Cheshire M33 7PE (GB).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): JOHNSON, Ian, David [GB/GB]; West Kingston Cottage, 23 Golden Avenue, East Preston, Littlehampton, Sussex BN16 1QY (GB). PIEKARSKI, Marek, Stephen [GB/GB]; 20 Gawsworth Road, Macclesfield, Cheshire SK11 8UE (GB).
- Published:
— with international search report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: DATA SWITCHING ARBITRATION ARRANGEMENTS



(57) Abstract: There is disclosed a masking unit (REQMSK) for use in a data packet switching system. The data switching system being of the type having a memoryless cross-back switch (SM) providing cyclic connections under the control of a switch arbiter (SCARB) between ingress routers (IR0, IR1, IR2 and IR3) and egress routers (ER0, ER1, ER2 and ER3). Each of the ingress routers (IR0 - IR3) is provided with incoming packet buffering on a virtual output queue basis (VOQ0.0, VOQ0.1, VOQ0.2, and VOQ0.3 for ingress router IR0). Each virtual output queue also produces a connection request signal REQ0.0 to REQ3.3 when the corresponding queue has a data packet in it. The masking unit REQMSK is arranged to randomly mask out correlated connection requests.

WO 01/76157 A1

DATA SWITCHING ARBITRATION ARRANGEMENTS

The present invention relates to data switching systems and is more particularly concerned with switch connection request arbitration arrangements for providing traffic flow control management for such systems.

In a data switching system of the type disclosed in PCT/GB99/03748 a central memoryless cross-bar type switch provides connections on a cyclic basis between ingress and egress routers. The ingress routers provide buffering for incoming data packets on a virtual output queue basis. Unless the data switching system connection arbitration arrangements for selecting the connections across the switch are controlled on a fair basis, severe congestion at the ingress virtual output queues can occur. For example, in a simple four ingress four egress router system if a traffic pattern occurs where the four ingress routers require connection to the same egress router for say 33% of the traffic cycle with the rest of their offered load distributed across the remaining egress routers, it can be shown that, without bandwidth allocation arrangements, all of the virtual output queues for that egress router will fill rapidly with an offered load of greater than 80% since the egress router is over-subscribed.

It is an object of the present invention to provide a connection request control mechanism that overcomes the above mentioned problem.

According to the invention there is provided a masking unit, for use in a data packet switching system of the type having a memoryless cross-bar type switch providing cyclic connections between ingress routers and egress routers, the ingress routers providing incoming packet buffering on an virtual output queue basis and generating switch connection requests when a virtual output queue contains a data

packet, characterised in that the masking unit receives all the switch connection requests and is arranged to randomly mask connection requests.

Each connection request is typically accompanied by a four-bit weight value. The weight value is set by the system parameters which defines a quality of service rating for the request. The quality of service rating may reflect the priority level of the virtual output queue and may be derived from the urgency required for the data packet to be switched through the system, the number of packet cycles elapsed since the buffer was served and the size of the buffer queue for example.

According to a feature of the invention the masking unit includes a comparator for each connection request and the comparator is fed with the weight of the request and a randomly generated bit stream, the comparator producing a random bit stream whose proportion of 1's to 0's is determined by the weighting factor of the request and the random bit stream is used for masking out the request.

According to a further feature of the invention the masking unit includes a two input AND gate for each request to which is connected on one input the connection request and on the other input the output of the comparator.

According to a further feature of the invention there is provided an arrangement for generating a uncorrelated set of randomly generated bit streams comprising a single random bit stream generator clocked at frequency f and fed into the first stage of a shift register which is n bits wide, where n equals the number of bits of the weighting factor of the requests, and $x+1$ stages deep, where x equals the maximum number of connections that the cross-bar switch can make and the shift register is clocked at a frequency for f/n to produce at each of the stages bar the first an uncorrelated set of bit streams for application to the comparators of the masking

unit.

The invention together with its various features will be more readily understood from the following description which should be read in conjunction with the accompanying drawings. Of the drawings:-

Figure 1 shows in block diagram form a data packet switching system according to the invention,

Figure 2 shows in abbreviated form the masking unit of the invention whereas, Figure 3 shows the circuit arrangement for generating nibble streams for use in Figure 2 and for generating uncorrelated masking of requests.

The basic architecture of a high speed data packet switching system using a memoryless space division (cross-bar) type switching matrix SM is shown in Figure 1. The switching system is shown for simplicity of presentation as a four ingress, four egress router arrangement. The ingress routers, of which only ingress routers IR0 and IR3 are shown in any detail, provide so-called virtual output buffering for ports IP0 to IP3 which are connected as required by a NxN (i.e. four by four) switching matrix SM which serves four egress routers ER0 to ER3 attached to four egress ports EP0 to EP3. Each ingress router includes virtual output queue buffers, such as VOQ0.0 to VOQ0.3 for ingress port IP0. Considering ingress router IR0 for ingress port IP0, buffer VOQ0.0 is the virtual output queue for egress router ER0, VOQ0.1 for egress router ER1, VOQ0.2 for egress router ER2 and VOQ0.3 for egress router ER3. Similarly ingress port IP3 is provided with four virtual output queues VOQ3.0 to VOQ3.3 on a similar basis. The switching matrix SM is controlled by a switch control arbitrator SCARB which analyses packet transfer requests REQ0.0 to REQ3.3 and determines for each switching matrix cycle the connections to be made from ingress port to egress port and,

therefore, the virtual output queues to be involved in these connections. The switch control arbiter SCARB also selects the virtual output queue within an ingress router to be used in a connection by activating an output selector SEC0 - SEC3.

The connection requests generated by the ingress ports, such as REQ0.0 for ingress port IP1 virtual output queue VOQ0.0, identify those virtual output queues that are storing packets of data and these are applied to a request mask arrangement REQMSK. For ease of presentation for example only the connection request paths from virtual output queues VOQ0.0 and VOQ0.3 for ingress router IR0 are shown in the drawing. It will be appreciated by those skilled in the art that each virtual output queue in each of the ingress routers has an individual connection request path to the request mask arrangement REQMSK. The request mask arrangement REQMSK will receive a request lead from each virtual output queue from each of the ingress routers (IR0 - IR3) making, in the four ingress router/four egress router system of Figure 1, a total of sixteen request lines REQ0.0 to REQ3.3.

The request mask arrangement REQMSK is introduced to solve the problem created by the switch control arbitrator SCARB when traffic conditions are such that congestion occurs at specific ingress port virtual output queues.

The solution to the congestion problem, provided by the request masking arrangement REQMSK, is to create a bandwidth allocation mechanism that allows the egress bandwidth to be randomly preferentially allocated to the data stream emanating from a port so that it appears unimpeded by the data streams emanating from the remaining ports. The implementation of the bandwidth allocation mechanism provided by the request masking arrangement REQMSK is based upon pseudo random probabilistic masking. That is any particular connection which has its bandwidth restricted to say 33% will have its request randomly masked from the

switch control arbitrator SCARB 67% of the time.

The request masking arrangement is shown in Figure 2 for two ingress port connection requests only (i.e. ingress port 0 and ingress port 1) for ease of presentation. Each virtual output queue request, such as REQ0.0, is applied to a two input AND gate such as G0.0 whose other input is fed from a comparator, such as C0.0. The comparator is used to provide a pseudo random bit stream whose proportion of 1's to 0's is determined by a randomly generated weight value, called a nibble stream ns, compared with the weighting factor for the request. Each connection request is typically accompanied by a four bit weight factor value, such as Wt (0.0) for request REQ0.0. The weight factor value is set by the system parameters to define the quality of service rating for the request (i.e. the virtual output queue). The quality of service rating may reflect the type of traffic handled by the egress port, the priority level of the egress port as well as the occupancy of the virtual output queue. The weighting factor value may be adjusted dynamically by the data switching system administration control and the implications of the weight factor value are outside the scope of this invention.

The randomly generated weight value is shown in Figure 2 as NS(0) for example and Figure 3 shows how the so called nibble streams are generated.

It is important to ensure that there is no interaction between the masking of any pair of connections (i.e. the "random" masking of requests must be uncorrelated in any pair of related connections). Related connections are defined by having either a common source or a common destination port. For example, when two ingress ports have 100% of their offered load for the same egress port, if the masking bit streams become correlated (i.e. both requests are masked out), there will be a period where no connections are made and there is a waste of possible

bandwidth. In addition correlation between bit streams can also lead to unfairness caused by the sequences that the requests are presented to the switch control arbiter SCARB. To overcome this effect the sixteen possible requests in a four port system are divided into 4 sets of unrelated connections and each of the sets is defined by a skew factor K where K is equal to the destination port address minus the source port address. When $K=0$ ingress port 0 can be connected to egress port 0, ingress port 1 to egress port 1, ingress port 2 to egress port 2 and ingress port 3 to egress port 3, whereas, when $K=1$, ingress port 0 can be connected to egress port 1, ingress port 1 to egress port 2, ingress port 2 to egress port 3 and ingress port 3 to egress port 0. Similarly when $K=2$, ingress port 0 can be connected to egress port 2, ingress port 1 to egress port 3, ingress port 2 to egress port 0 and ingress port 3 to egress port 1 whereas when $K=3$, ingress port 0 can be connected to egress port 3, ingress port 1 to egress port 0, ingress port 2 to egress port 1 and ingress port 3 to egress port 2. With such an arrangement the problem of generating multiple uncorrelated bit-streams is reduced to generating one bit-stream per port (or skew factor) instead of one per connection. Rather than generating the set of bit streams for each skew factor, it is more efficient to generate a set of uncorrelated nibble streams and then generate the masking bit streams for each skew factor.

The nibble stream generator is shown in Figure 3. The nibble stream generator consists of a linear feed back shift register LFSR which generates a single pseudo-random bit-stream which is fed into an n bit wide nibble register NR at a frequency of f . Typically the frequency f will be the frequency of the clock cycle for the switching matrix SM multiplied by n . The n bit wide nibble in register NR is then shifted in parallel at a frequency f/n (i.e. the clock cycle frequency of the switching matrix SM) through a set of n -bit wide shift register stages NSR(0) - NSR(N) depending upon the number of cross-bar switch ports (i.e. four in the example chosen for the specific description). The linear feed back shift register

LFSR is clocked n times faster than the nibble shift register to ensure that successive nibbles are uncorrelated, where n is the width of the nibble (i.e. the number of bits in a weight value). The output from the nibble generator on leads NS(0), NS(1), NS(2) and NS(3) provides a random uncorrelated four bit weight for each skew factor. The randomly generated weighted nibble streams are applied to the comparators C0.0 etc. in the request masking arrangement of Figure 2. It will be noted that the effect of the skew factor is achieved by slipping the nibble streams in the same manner as that discussed for the uncorrelated requests so that for ingress port IP0 nibble stream NS(0) is applied to comparator C0.0 but to comparator C1.1 for ingress port IP1 and accordingly for nibble streams NS(2) and NS(3) as well as for ingress ports IP2 and IP3.

The invention has been described with reference to a simplified 4x4 ingress/egress router with weighted requests using a four bit weight value and it will be appreciated by those skilled in the art that the arrangements described can be adapted to accommodate more than four ingress/egress router arrangements and to use weight values requiring more or less bits.

CLAIMS

1. A masking unit (REMSK) for use in a data packet switching system of the type having a memoryless cross-bar switch (SM) providing cyclic connections between ingress routers (IR0, IR1, IR2, IR3) and egress routers (ER0, ER1, ER2, ER3), the ingress routers providing incoming packet buffering on a virtual output queue basis and being arranged to generate switch connection requests when a virtual output queue (VOQ) contains a data packet, characterised in that the masking unit (REMSK) is arranged to receive all of the switch connection requests (REQ) and to randomly mask connection requests (REQ).
2. A masking unit (REMSK) as claimed in claim 1, the masking unit including a comparator (C) for each virtual output queue, the masking unit being arranged to receive with each request an associated weight value (Wt), and to feed the weight values to the corresponding comparators (C) together with a stream of randomly generated values, the comparators (C) being arranged to produce respective random bit streams whose proportion of 1's to 0's is determined by the corresponding weight values (Wt), and the masking unit being arranged to use each random bit stream for masking out the requests (REQ) from the corresponding virtual output queue (VOQ).
3. A masking unit (REMSK) as claimed in claim 2 further comprising a two input AND gate (G) for each comparator (C), one input of the AND gate (G) being arranged to receive the corresponding connection requests and the other input of the AND gate (G) being arranged to receive the output of the corresponding comparator (C).

4. A masking unit as claimed in claim 2 or 3 including an arrangement for generating an uncorrelated set of randomly generated bit streams, the arrangement comprising a single random bit stream generator clocked at frequency f and fed into the first stage of a shift register (NR) which is n bits wide, where n equals the number of bits of the weighting factor of the requests, and $x+1$ stages deep, where x equals the number of connections that the memoryless cross-bar switch can make, the shift register (NR) being clocked at a frequency f/n to produce at each of the stages (Ns) bar the first an uncorrelated set of n bit nibble streams for application to the comparators (C) of the masking unit.

5. A masking unit as claimed in claim 4 in which the frequency f/n is equal to the frequency of the cyclic operation of the memoryless cross-bar switch (SM).

6. A masking unit as claimed in claim 4 or 5 in which the random bit stream generator is a linear feed back register (LFSR).

7. A method of controlling a data packet switching system of the type having a memoryless cross-bar switch (SM) providing cyclic connections between ingress routers (IR0, IR1, IR2, IR3) and egress routers (ER0, ER1, ER2, ER3) under the control of a switch control arbiter (SCARB), the ingress routers providing incoming packet buffering on a virtual output queue basis and being arranged to generate switch correction requests (REQ) when a virtual output queue (VOQ) contains a data packet, the method being characterised by comprising:

randomly selecting ones of said switch connection requests; and

transmitting all the connection requests (REQ) except said selected requests to the switch control arbiter (SCARB).

8. A method according to claim 7 in which each request (REQ) is associated with a weight value (Wt), the selection of the requests including comparing the associated weight value (Wt) with a first randomly generated value to generate a second random value and selecting the requests based on the second random value.

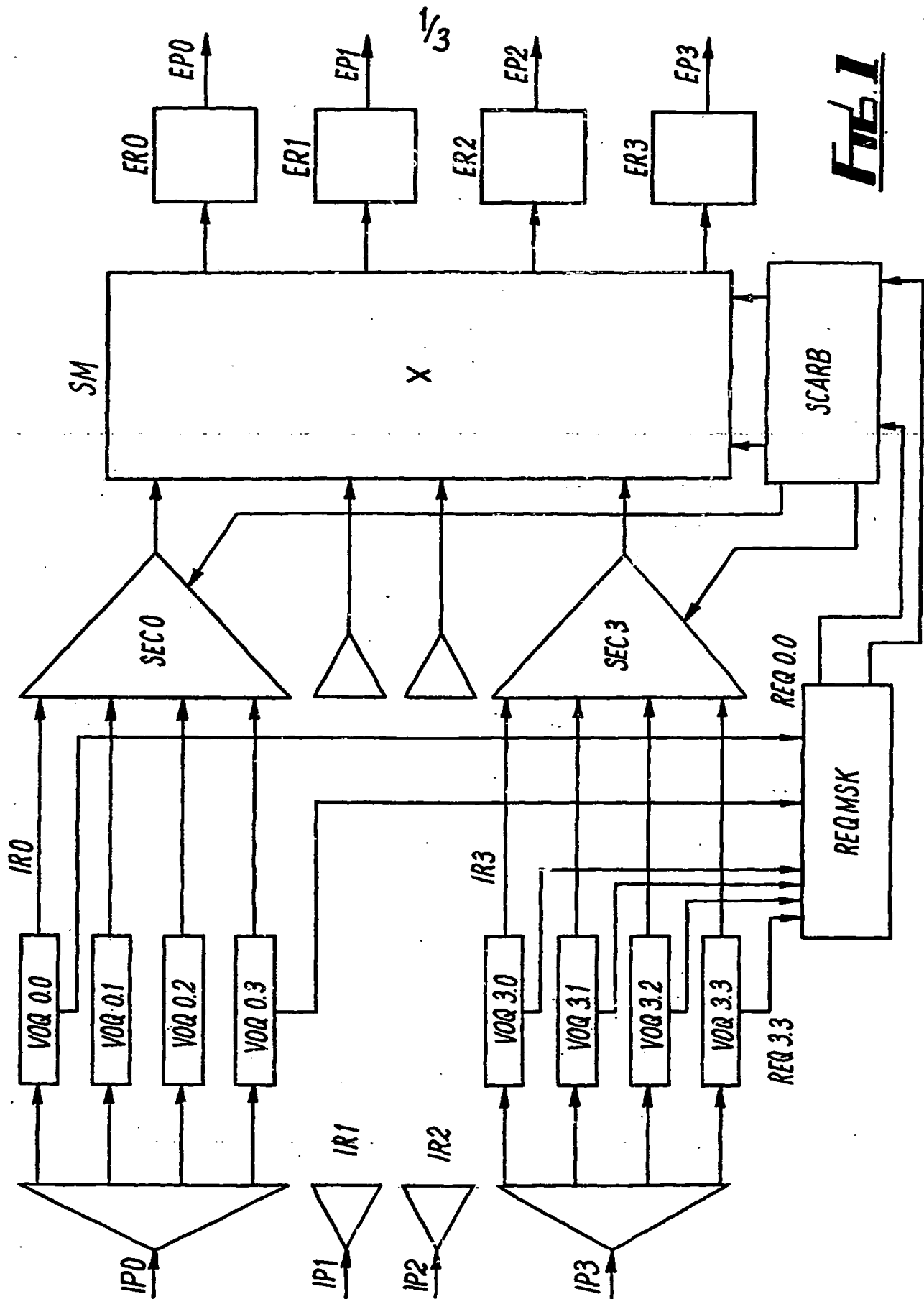
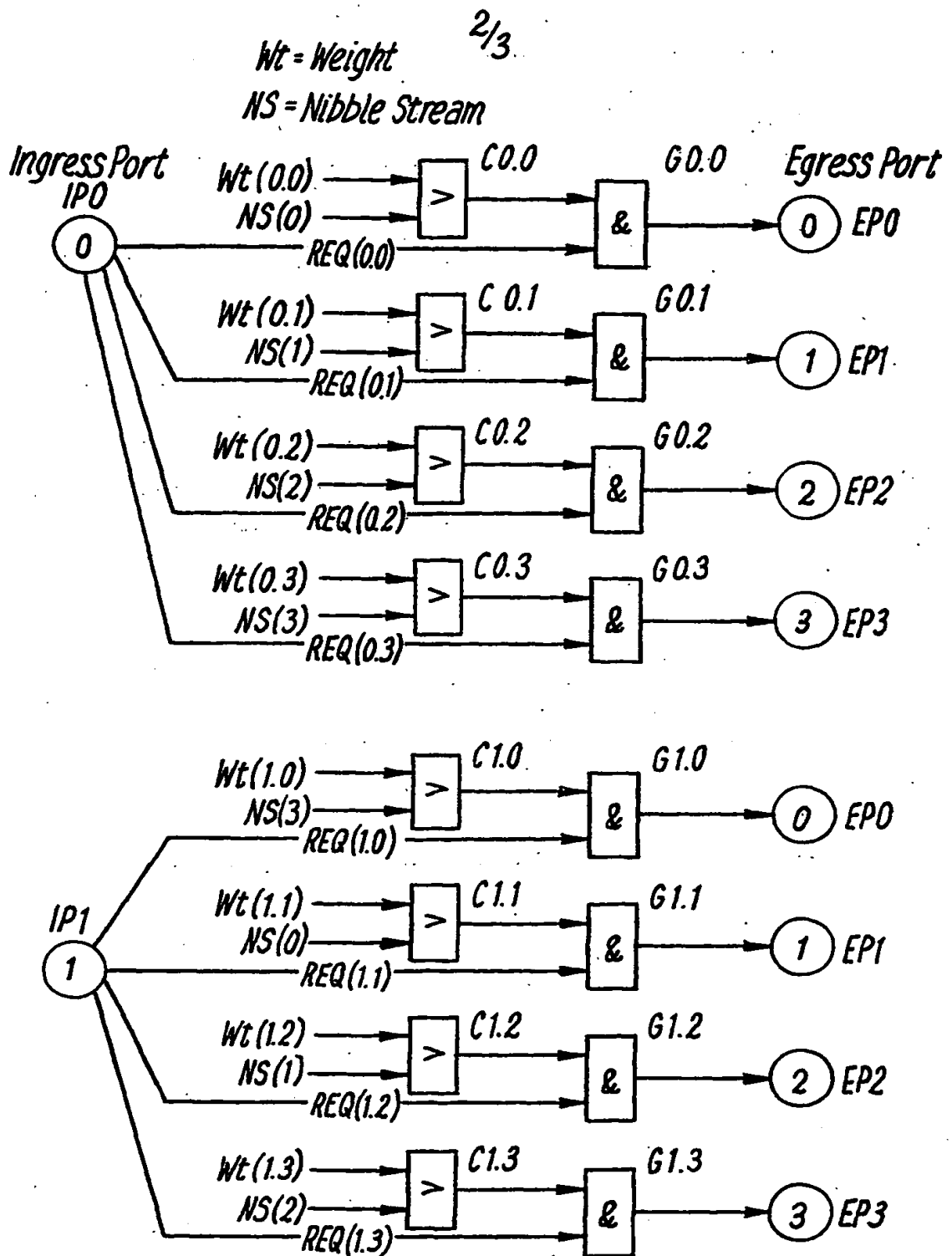
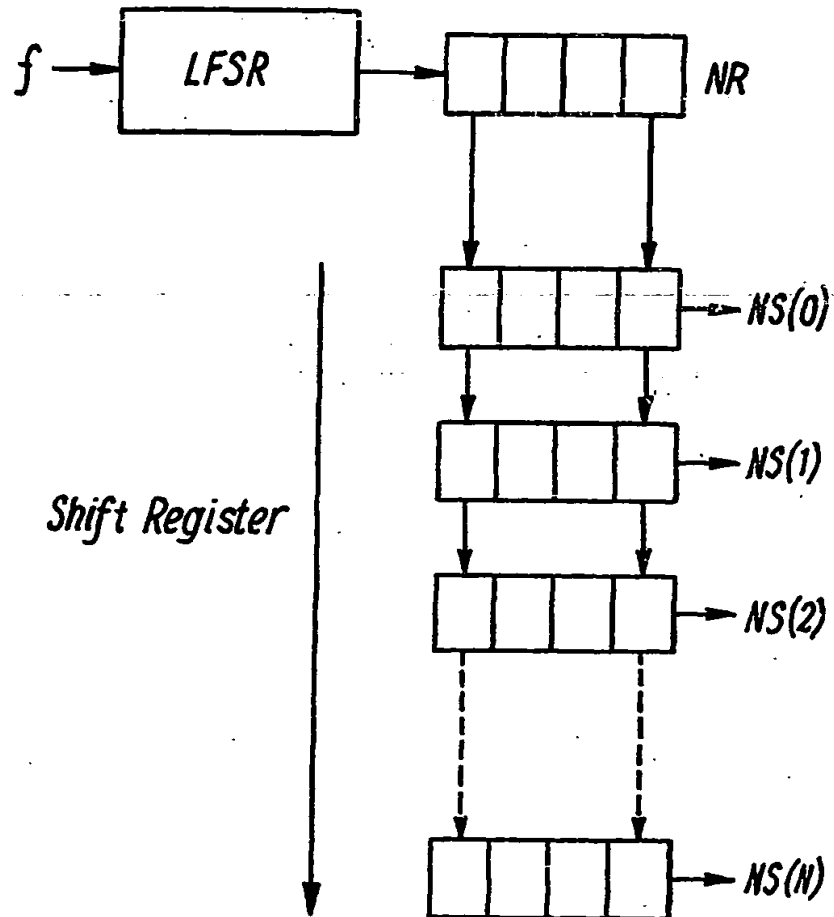


FIG. 1

**FIG. 2**

3/3

**FIG. 3**

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 01/01073

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 H04L12/56

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6 044 061 A (AYBAY GUENES ET AL) 28 March 2000 (2000-03-28)	1,7
A	column 2, line 65 -column 4, line 65; figure 8	2-6,8
A	US 5 541 916 A (NOBUYUKI OBA) 30 July 1996 (1996-07-30)	1-8
	column 6, line 24 -column 7, line 2; figure 5	
A	US 5 500 858 A (MCKEOWN NICHOLAS W) 19 March 1996 (1996-03-19)	1-8
	column 3, line 10 - line 38; figure 6	
A	US 5 267 235 A (THACKER CHARLES P) 30 November 1993 (1993-11-30)	1-8
	figures 2-4	



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

Z document member of the same patent family

Date of the actual completion of the international search

7 June 2001

Date of mailing of the international search report

22/06/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Meurisse, W

INTERNATIONAL SEARCH REPORT

Information on patent family members

Intel nal Application No
PCT/GB 01/01073

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 6044061	A	28-03-2000	AU 2980099 A EP 0981878 A WO 9946903 A	27-09-1999 01-03-2000 16-09-1999
US 5541916	A	30-07-1996	JP 2502920 B JP 7087094 A	29-05-1996 31-03-1995
US 5500858	A	19-03-1996	NONE	
US 5267235	A	30-11-1993	NONE	